# Application of Genetic Algorithm and Support Vector Machine in Classification of Multisource Remote Sensing Data

H. T. Chu, L. Ge*, A. H.-M Ng, C. Rizos

School of Surveying and Geospatial Engineering, the University of New South Wales,
Sydney NSW 2052, Australia
l.ge@unsw.edu.au

*Abstract*-**The use of multi-source remote sensing data for improved land cover classification has attracted the attention of many researchers. On the other hand, such an approach increases the data volume with more redundant information and increased levels of uncertainty within datasets, which may actually reduce the classification accuracy. It is therefore an essential, though challenging task to select appropriate features and combine datasets for classification. The combination of feature selection techniques using the Genetic Algorithm (GA) and Support Vector Machines (SVMs) classifiers has been used in various application fields in a number of studies on classification of hyperspectral data. However, the performance of this technique for classifying multi-source remote sensing data has not been well evaluated in the literature. In this study, the GA-SVM model was proposed and implemented to classify multiple combined datasets, consisting of Landsat 5 TM, multi-date dual polarization ALOS/PALSAR images and their multi-scale textural information. The performance of the proposed method was compared with that of the traditional stack-vector approach. A large number of different combined datasets were generated and classified. It is revealed that the proposed method is very efficient for handling multisource data. Results indicated that the GA-SVM approach clearly outperforms the stack-vector approach, with significantly higher classification accuracy and much fewer input features. The highest classification accuracy achieved was 96.47% with only 81 out of 189 features being selected. This study also demonstrated the advantages of using multi-source data over single source data.**

***Keywords- Multi-Source Data; Support Vector Machine; Genetic Algorithm; SAR; Optical Imagery; Texture***

## I. INTRODUCTION

### *A. Uses of Multi-Source Remote Sensing Data for Land Cover Classification*

Land cover classification is one of the most important applications of remote sensing. The advantages of remote sensing for providing objective data with large areal coverage, over multiple revisit dates, and with a diversity of imaged characteristics from a large variety of sensors are well recognised and make it very suitable for mapping land cover features. Remotely sensed data, in particular satellite imagery, can be acquired in various regions of the electromagnetic spectrum, from the visible-near infrared (optical) to the microwave (radar) parts of the spectrum. Consequently, different kinds of satellite imagery detect different characteristics of ground surfaces. For instances, optical images from missions such as Landsat, SPOT, MODIS, IKONOS or QuickBird provide information essentially on reflectance and absorption capability of land cover features, since the imaging sensors are sensitive to the visible to near-infrared regions of the spectrum. On the other hand, Synthetic Aperture Radar (SAR) imagery, provided by missions such as ENVISAT/ASAR, ALOS/PALSAR or TerraSAR-X, sensitive to the microwave region of the spectrum, contains information on surface roughness, dielectric content and the structures of the illuminated ground or vegetation

Thus a combination of SAR and optical images could provide complimentary information and lead to improved land cover classification results. Numerous studies using this integration approach have been reported, with different datasets and using different classification techniques [1], [2], [3], [4], [5], [6], [7]. Although the reported results vary considerably, most authors claim that the integration of SAR and optical data has improved classification performance.

In [5] land covers were classified using several combinations of Landsat ETM+ and Radarsat images. The overall classification accuracy using combined datasets was improved to 74.60% as compared to an accuracy of 69.35% using only Landsat ETM+ data. The synergy of dual-polarimetric SAR (ENVISAT/ASAR) satellite image data and optical medium resolution (Landsat ETM+) data for land cover classification at the regional level in a test site in Central Sulawesi, Indonesia has been investigated in [2]. The authors pointed out that the integration of ASAR with Landsat images increased classification accuracy significantly, with the combination of like-polarised ASAR time series and Landsat multi-spectral data producing the best results. In [6] classifications of various land cover features in the south of Vietnam were investigated using a combination of multi-temporal ALOS/PALSAR (L-band), ENVISAT/ASAR (C-band) SAR and SPOT multi-spectral optical satellite data. Results demonstrate the advantages of the integration approach and clearly highlighted the complementary nature of multi-source datasets. The combination of optical and multi-temporal SAR images has resulted in remarkable improvements in classification accuracy of 6.45% and 23.13% (SPOT + EtNVISAT/ASAR) and 10.01% and 29.4% (SPOT + ALOS/PALSAR) in comparison to the cases of using only SPOT 4 multi-spectral, ASAR or PALSAR multi-date images individually. It was

also reported that the combined utilisation of optical (Landsat ETM+) and SAR (Radarsat-1) provided useful information on land cover classes and improved classification accuracy compared with using either type of original image data [7].

Apart from the spectral (optical) and backscattered (SAR) information within satellite imagery, image textural information also plays a crucial role in land cover classification. Image texture provides information on spatial arrangement and variation of patterns on the earth's surface. While the traditional approach using spectral or backscattered data might not be sufficient for land cover mapping due to problems of similarity between classes and variation within class, the textural information is considered to be useful additional data. Textural information has been combined with spectral or backscattered data for land cover classification [8], [9]. The investigation of combining Landsat ETM+ and Radarsat SAR data with textures for classification over the Brazilian Amazon Basin area was conducted in [10]. Texture information derived from the grey-level co-occurrence matrix (GLCM) based on information from Landsat ETM+ panchromatic band and Radarsat data using different sized moving windows were examined. These textures were combined with original data using a maximum likelihood classification process. Results of classification demonstrated improved overall accuracies by 5.8% to 6.9% compared to the classification based on Landsat ETM+ data.

The use of spectral/textural classification schema was carried out in [11]. In this study, textural information, including GLCM and edge-density measures generated from an IKONOS image, was incorporated with spectral data. Results showed that the spectral/textural approach obtained an overall accuracy of 80.01% compared to 63.44% when using only spectral bands. In [12], the multi-scale texture approach was applied. The first-order (variance) and second-order (GLCM entropy) texture measures derived from different window sizes were employed as additional information for forest stand classification. All of the different texture measures provided improvements in overall accuracy from 4 to 13%. The multi-scale image texture approach caused significant increases of 4 to 8% compared with using single-band texture measures. Results of this study also indicated that there was no single window size that could sufficiently represent the whole range of textural information in the image. The robustness of the multi-scale texture analysis was also mentioned by [13], who applied GLCM texture measures derived from different window sizes and high spatial resolution IKONOS imagery for urban land cover/use classification. According to [13], the overall accuracy of the multi-scale approach was higher than cases of single-scale texture and original spectral data by ~6% and ~11%, respectively. However, adding textural information does not always increase the classification accuracy. It is claimed that radar texture did not give any improvement instead reduced the overall accuracy for some classes [5]. In [6], incorporation of textural information extracted from either optical or SAR images did not give any significant improvement.

Nevertheless, integration of multi-source remotely sensed optical and SAR imagery and image textural data has the potential to improve the results of land cover mapping since additional information is used in the classification process. On the other hand, this integration increases the data volume, but with large amounts of highly correlated features and redundant information. Unfortunately, a large data volume on its own does not necessarily lead to an increase in classification accuracy. According to [14], the mean accuracy will increase until it reaches the peak value, beyond which no significant improvement will be achieved with additional measurements. This is the so-called Hughes phenomenon. Lu & Weng [15] also stated that due to different capabilities in land cover separability, utilisation of too many input data for classification may not improve (but can actually decrease) the classification accuracy, and it is essential to select only input variables that are useful for discriminating land cover classes. Therefore, selecting optimally combined datasets which give the best classification accuracy is a challenging task.

### *B. Advantages of Support Vector Machine Classifier for Land Cover Mapping*

In addition to the input data, the employed classification techniques are also vital for land cover mapping. A broad range of classification algorithms has been developed and applied for classifying remotely sensed data. The traditional parametric classifiers such as the Maximum Likelihood (ML) classifier are commonly used [16], [17] because of its acceptable accuracy and fast performance. However, the major limitation of the ML algorithms is the assumption of normal distribution of input data – which is often not true for remotely sensed data [18]. This limitation makes it difficult for such parametric classifiers to handle complex datasets such as multi-source data. Unlike parametric classifiers, non-parametric classifiers such as the Support Vector Machine (SVM) do not constrain their application with the assumption of normal distribution, and are therefore often considered more appropriate for classifying remotely sensed data.

SVMs are a recent development of a non-parametric supervised classification technique, which have proven to be very robust and reliable in the field of machine learning and pattern recognition [18], [19]. In SVMs, the problem of over-fitting in classification of high dimension feature space is controlled by the structure risk minimisation principle. The SVMs have been applied successfully in many studies using remotely sensed imagery. In these studies the SVMs often provided better (or at least the same level of) accuracy as other classifiers [18]. Pal & Mather [20] compared SVMs, ML and the artificial neural network (ANN) approach for classifying Landsat 7 ETM+ and hyperspectral (DAIS) data. The results showed that SVMs obtained higher classification accuracy than either the ML or ANN classifier. Kavzoglu & Colkesen [21] used Terra ASTER images and SVMs with radial basis and a polynomial kernel function to classify land

cover type in the Gebze District of Turkey. The performance of SVMs was compared with the ML classifier. Results indicated that SVMs in most cases outperformed the ML algorithm in terms of overall accuracy (by 4%) and individual classes. It was also found that the radial basis function (RBF) kernel gave higher accuracy than the polynomial kernel by approximately 2% in overall accuracy. In [19], the SVMs classifier with different kernel functions including linear, radial, sigmoid and polynomial, was used to classify SPOT 5 satellite images. The performance of SVM classifiers was compared with the Decision Trees (DTs). Results showed that SVMs outperformed DTs in terms of classification accuracy. The lowest overall classification accuracy given by SVM classifiers was 73.70% with the linear function kernel while the highest accuracy of 76.00% was obtained by the radical basic function kernel. The overall classification accuracy of the DTs algorithm was only 68.78%.

In this study the SVM classifier with the RBF kernel was applied to classify land cover classes. Two parameters were optimally specified in order to ensure the best accuracy: the penalty parameter C and the width of the kernel function γ. The common way of finding the optimal C and γ is using a grid search algorithm [22].

### *C. Genetic Algorithm (GA) and SVMs*

In order to resolve the problem of finding the optimal input datasets for classification the Feature Selection (FS) techniques were applied. There are numerous FS techniques including exhaustive search, sequential feature selection (forward and backward), branch and bound, simulated annealing and the Genetic Algorithm (GA). Among the many FS techniques that have been used, the GA has been proven to be very effective for handling global optimisation problems with large datasets, and has less chances of converging to a local optimal solution than other methods [23], [24], [25]. The FS used an objectives function to evaluate candidate subsets and return measures of their 'goodness' [26]. There are two kinds of objective functions: filters and wrappers. The filter approach analyses feature subsets based on their information content, such as distance between classes (separability index) and statistical correlation. In the wrapper approach, the feature subsets are evaluated in relation to the classification accuracy. Consequently, the choice of objective function is dependent on the classifiers used. The main advantages of the filter approach are low computation cost and good generality. However, the filter method has a tendency to select the whole dataset as the optimal solution and often results in lower accuracy than the wrapper method. The wrapper method generally yields better accuracy than the filter method since the candidate dataset interacts directly with the specific classifier. The main limitation of the wrapper method is its rather slow computation. Nevertheless, with the availability of powerful computer systems, the wrapper approach has become more attractive to researchers. As mentioned previously, the accuracy and efficiency of the SVM classifier depends on both input datasets and the kernel parameters. While other methods can only deal with a single issue at a time, the GA techniques can find the optimal feature subset and kernel parameters at the same time.

The GA-SVM model has been applied very successfully in many applications, including biology, medical and financial data analysis (for example, [23], [24], [28]). In the field of remote sensing, few studies have been conducted for classification of hyperspectral data [24], [25], [27]. However, the application of GA-SVM model for classifying multi-source remotely sensed data has not been previously reported in the literature. Thus, the objectives of this study are: 1) to evaluate the integration of optical, SAR satellite images, and their textural information for land cover mapping; and 2) to propose and implement the combination of feature selection with GA techniques and SVMs for classifying multi-source remotely sensed data.

The paper is organised as follows. A brief introduction of the SVM classifier is given in Section 2. The basic concept of the GA is described in Section 3. Section 4 describes the study area and dataset, the methodology is presented in Section 5. Section 6 presents the results and discussion, and the conclusions are drawn in Section 7.

## II. SUPPORT VECTOR MACHINE

A SVM aims to separate two classes by determining an optimal hyper-plane that maximises the margin between these classes in a multi-dimensional feature space [21]. The optimal hyper-plane is determined by using only the closest training samples – namely the 'support vectors' in the training datasets. Hence, the approach only considers samples close to the class boundary and works well with small training sets, even when high-dimensional datasets are being classified.

As in a case of a binary classification, in n-dimensional feature space, $x_i$ is a training set of m samples, i=1, 2, …, m, and their class labels $y_i$ = -1 or +1.

The optimum separation plane is defined as:

$$w.x_i + b \le -1, \quad \text{as x belong to class 0} \tag{1}$$

$$w.x_i + b \ge +1, \quad \text{as x belong to class 1} \tag{2}$$

$$\text{or} \qquad y_i\left[w.x_i + b\right] \ge 1 \quad \forall i \tag{3}$$

In practice classes are not always fully separated by linear boundaries. Thus, the error variable $\xi_i$ is introduced. So, Equation (3) becomes:

$$y_i\left[w.x_i + b\right] \ge 1 - \xi_i, \quad \xi_i \ge 0 \tag{4}$$

The optimum hyper-plane is identified based on solving optimisation problems:

$$\min\left[\left(\tfrac{1}{2}\right)w^2 + C\sum_{i=1}^{m}\xi_i\right] \tag{5}$$

where C is the penalty parameter according to the error $\xi_i$.

For nonlinear classification the SVM projects input data into a higher dimensional space using a nonlinear vector

mapping function $\varphi$. In order to reduce the burden of computation, Vapnik [29] proposes a kernel function $K(x, y)$, in which: $K(x_i, y_j) = \varphi(x_i) \times \varphi(y_j)$.

Using the technique of Lagrange multipliers, the optimisation problem becomes:

$$\min \tfrac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{m}\alpha_i \quad (6)$$

$$with \sum_{i=1}^{m} y_i\alpha_i = 0 \quad and \quad 0 \le \alpha_i \le C, \ i=1, 2, \ldots, m$$

where α is the Lagrange multipliers. The Lagrangian has to be minimised with respect to $w$, $b$ and maximised with respect to $\alpha \geq 0$.

Major kernel functions are the Gaussian Radial Basis Function (RBF), linear, polynomial and sigmoid functions:

Linear $$K(x, y) = x.y \quad (7)$$

RBF $$K(x, y) = \exp\left(-\gamma\|x - y\|^2\right) \quad (8)$$

Polynomial $$K(x, y) = \left((x, y) + 1\right)^d \quad (9)$$

Sigmoidal $$K(x, y) = \tanh\left(k(x.y) + 1\right) \quad (10)$$

The final decision function is defined as:

$$f(x) = \text{sign}\left(\sum_{i=1}^{m} y_i\alpha_i K(x_i, x) + b\right) \quad (11)$$

The above theory was developed only for separating two classes. For the cases of multi-classes, several strategies have been proposed to apply SVMs. The most common approaches are one-against-all (OAA), and one-against-one (OAO). Let us assume there are N classes. In the one-against-all strategy a set of N binary SVM classifiers, each trained to separate one class from the rest, is applied. The pixel will be labelled with the class in which the pixel has the maximum decision value. On the other hand, in the one-against-one strategy N(N-1)/2 SVMs are constructed for each pair of classes. Each classifier votes to one class, and the pixel will be assigned to the class with the most votes. In this study the one-against-all strategy, which is widely used in the literature [16], [20], [21] was chosen for land cover/use classification.

## III. GENETIC ALGORITHM

The GA is a method for solving optimisation problems based on the concept of 'natural selection', which has its roots in biological processes. At the first stage the set of features are generated randomly as a population. In later steps, the GA selects individuals as 'parents' from the current population according to the values of fitness function, and produces 'children' for the next generation. Over successive generations the GA modifies the population toward an optimal solution based on fitness functions and operations such as crossover and mutation. The GA can work with a large number of features and is considered an efficient method for feature selection [23], [24], [27].

There are three main operations employed in the GA, namely selection, crossover and mutation. The selection operator selects the individuals to be 'parents' that will help to reproduce the population in the next generation. The crossover operator combines two 'parents' to generate new individuals for the new generation. The mutation operator randomly changes 'parents' to generate new 'children'. Figure 1 below illustrates the crossover and mutation operator in GA.

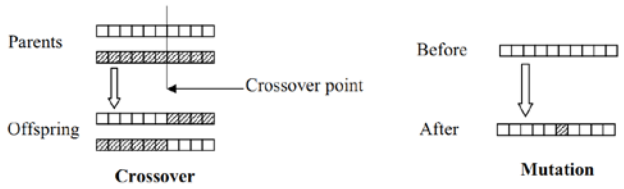


Figure 1 Illustration of the crossover and mutation operators [23]

The GA model for feature selection and parameter optimisation involves designing of chromosomes, definition of the fitness function and architecture of the system.

### A. Chromosome Design

As the SVM with Radial Basic Function (RBF) kernel was applied to classify land cover features in the test site, the two parameters C and γ are to be defined. The GA-based model will try to optimise both input features and the SVM's parameters. Thus, the chromosome consisted of three parts, representing selected features, C and γ. A binary coding technique was used to define the chromosome [23]. In Figure 2, $Ib_1$ to $Ib_{Nf}$ represent input features, $Ib_i$=1 means a corresponding feature is selected, $Ib_i$=0 means a feature is not selected. $Cb_1$ ~ $Cb_{Nc}$ represents the value of C and $\gamma b_1$ ~ $\gamma b_{N\gamma}$ represents the value of γ.

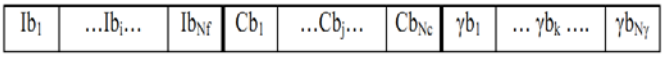


Figure 2 The binary coding of the chromosome

### B. Fitness Function

The fitness function is designed to test whether an individual is 'fit' for reproduction processes. The chromosomes that have the higher fitness value will have more chances to be chosen as parents or selected for recombination in the next generation. In previous studies, two criteria are often used for designing the fitness functions: classification accuracy and the number of selected features [23], [24], [25], [27]. In this paper, we proposed a modified design of the fitness function with an additional criterion, namely, average correlation.

$$Fitness = W_{\text{OA}} \times \frac{100}{OA_{\text{SVM}}} + W_{\text{f}} \times \frac{Ns}{N} \times Cor \quad (12)$$

Where $OA_{SVM}$ is the overall classification accuracy (%), $W_{OA}$ represents the weight for the classification accuracy, $W_f$ represents the weight for the number of selected features, $N_S$ is the number of selected features, and $N$ is the total number of input features. $Cor$ is the average correlation coefficient

of selected bands. The values of $W_{OA}$ and $W_f$ can be set differently based on user requirement.

The major steps for GA-SVM feature and parameter selection are (Figure 3):

1) Randomly create chromosomes of the initial population. The size of the initial population should be selected by users.

2) Calculate the fitness value of each individual in the population. This step involves converting the binary code of the chromosomes to identify C, γ and the selected features. The SVM classifier will implement the classification based on these values and the training datasets. The fitness value of the individual is calculated by Equation (12).

3) In the reproduction step a number of individuals with a high fitness value will be selected and kept for the next generation. The other individuals will be used for the crossover and mutation process to generate new children for the next generation.

4) If the stopping condition is satisfied the evolution terminates and the optimal result represented by the best individual is returned, otherwise the evolution will continue. The stopping criterion is usually a predefined change of the fitness values or maximum number of generations.

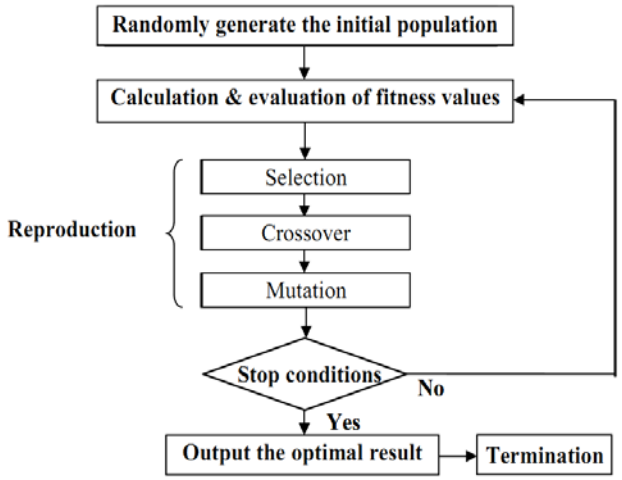


Figure 3 Major steps of feature selection using the Genetic Algorithm

## IV. STUDY AREA AND DATA

The study area was located in Western Australia (WA), Australia, with the centre coordinate being 116°57'45''E, 33°48'40''S. The site is characterised by relatively flat terrain with pastures, crops, sparse and dense tree cover. There are also some small rural residential settlements in the south of the study area. Two kinds of satellite imagery were employed for this study (Figure 4):

Synthetic Aperture Radar (SAR): 4 ALOS/PALSAR HH/HV dual-polarisation images acquired in 2010 (Table 1).

Optical: Landsat 5 TM image acquired on 07/10/2010 with spatial resolution of 30m. In this study, 5 spectral bands (from 1 to 5) were used.

TABLE I ALOS/PALSAR IMAGES FOR THE STUDY AREA

| Satellite/Sensor | Path | Acquisition Dates | Polarisation | Orbit |
|---|---|---|---|---|
| ALOS/PALSAR | 433 | 20/07/2010 | HH/HV | Ascending |
| | | 04/09/2010 | HH/HV | Ascending |
| | | 20/10/2010 | HH/HV | Ascending |
| | | 05/12/2010 | HH/HV | Ascending |

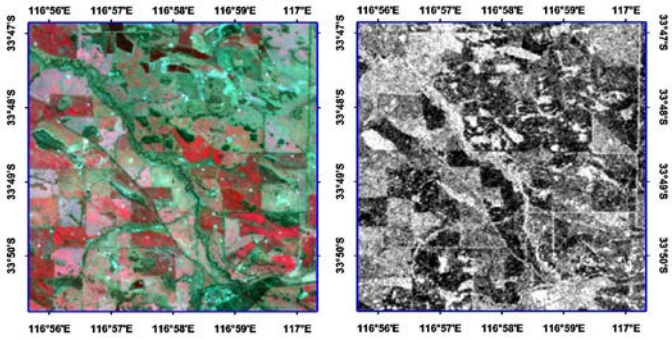


Figure 4 Landsat 5 TM (left) and ALOS/PALSAR HH (right) over the study area acquired on 07/10 and 20/10/2010, respectively

## V. METHODOLOGY

### A. Data Processing

ALOS/PALSAR and Landsat 5 TM images were registered to the map coordinate system (UTM projection, WGS84 datum) and resampled to 10m pixel size. Speckle noise in the PALSAR images was filtered by the Enhanced Lee Filter [30] with a 5x5 window size. SAR backscattered values were converted to decibel (dB) using:

$$Db = 10 \times \log_{10}(DN^2) \quad (13)$$

where Db, DN are magnitude values.

Besides the original ALOS/PALSAR images, several derivative images were also generated and integrated for classification, including the Temporal Backscattered Change, average and dual-polarised SAR index images. The Temporal Backscattered Change ($\mathbf{SAR}_{CH}$) image was generated based on all four different SAR images. This image highlights differences between the stable or non-changing features (such as urban areas, permanent vegetation, or still water) and temporarily changing features (such as annual crops):

$$\mathbf{SAR}_{CH} = \mathrm{Max}(Db_1,...,Db_n) - \mathrm{Min}(Db_1,...,Db_n) \quad (14)$$

where $\mathbf{SAR}_{CH}$ is the Temporal Backscattered Change image, the corresponding change images for HH and HV polarised images are $\mathbf{HH}_{CH}$ and $\mathbf{HV}_{CH}$. $Db_1$, $Db_2$, … $Db_n$ are backscattered values of pixels in corresponding SAR images, and the Max and Min denote the functions to pick up the largest and smallest pixel values within all applied images.

$$\mathbf{SAR}_{ind} = \frac{HH - HV}{HH + HV} \quad (15)$$

where $\mathbf{SAR}_{ind}$ is the SAR index image derived from the corresponding SAR dual HH/HV polarised image.

Two groups of texture measures were extracted and employed for classification, namely first-order and second-order texture measures. The first-order texture measures involve statistics computed directly from the original image and do not explain relationships with neighbouring pixels. On the other hand, the second-order texture measures consider the mutual dependence of sets of surrounding pixels [12], [31]. The most widely used second-order textural data are the grey-level co-occurrence matrix (GLCM), which measures relationships between pairs of pixels within a neighborhood. The First Principal Components (PC1) images generated from each of the four SAR and Landsat 5 TM images were used to derive textural information. In order to reduce correlation within datasets, it was necessary to select only components which are less correlated to each other. Hence only three first-order texture measures, namely Mean, Variance and Data range, and four GLCM texture measures, including Variance, Homogeneity, Entropy and Correlation, were employed. Since there is no preferred direction, the GLCM texture measures were computed as average of texture measures generated for eight different directions of $0^o$, $45^o$, $90^o$, $135^o$, $180^o$, $225^o$, $270^o$, $315^o$. Texture measures were calculated using Equations 16 to 22 below. As the multi-scale texture approach was adopted, textural data were generated from eight window sizes, including 3x3, 5x5, 7x7, 9x9, 11x11, 13x13, 15x15, 17x17, and used simultaneously.

First-order texture measures (F_OR):

$$Mean = \frac{1}{W}\sum_{k=0}^{n} i \times f_i \quad (16)$$

$$Variance = \frac{1}{W}\sum_{i=0}^{n} (i - \mathrm{Ave})^2 \times f_i \quad (17)$$

$$Data\_Range = \mathrm{Max}(i) - \mathrm{Min}(i) \quad (18)$$

where $f_i$ is number of pixel's value $i$ appeared and $W$ is the total number of pixels in a moving windows, $n$ is quantization levels of digital images [33].

Second-order texture measures (GLCM):

GLCM Variance $$\sigma_i^{\,2} = \sum_{i,j=0}^{N-1} P_{i,j}(i - \mu_i)^2 \quad (19)$$

GLCM Homogeneity $$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i-j)^2} \quad (20)$$

GLCM Entropy $$\sum_{i,j=0}^{N-1} P_{i,j} \times (-\ln P_{i,j}) \quad (21)$$

GLCM Correlation

$$\sum_{i,j=0}^{N-1} P_{i,j} \left[ \frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] \quad (22)$$

where $i$, $j$ are pixel's grey values; $P_{ij}$ is number of the co-occurrence of grey values $i$ and $j$, and N is the size of the moving window [31].

### *B. Integration, Feature Selection, Parameters Optimization and Classification*

Different combined datasets were generated, including Landsat 5 TM + its textures, SAR single- and dual-polarised images + their textures, SAR dual-polarised images + intermediate derived images, Landsat 5 TM + SAR single-/dual-images, Landsat 5 TM + SAR single-/dual-images + textures and intermediate images. These integrated datasets were classified using the SVM classifier with the RBF kernel. Two approaches, namely stack-vector and feature-selection using GA were implemented and the results were compared. The stack-vector is the most straightforward approach, where the data are added together as input in the classification process. In the feature-selection approach, data are selected to form datasets, which give the best solution based on the GA. The stack-vector approach was applied for all datasets including the original images and combinations. The feature-selection approach was only applied for the complex datasets (more than 12 input features), where the GA techniques were applied in order to optimise input data and the SVM's parameters at the same time.

The chromosome and the fitness function for the GA were designed as shown in Figure 2, and Equation 12. In this study the weight for classification accuracy ($W_{OA}$) and the weight for the number of selected features ($W_f$) were set within 0.65-0.8 and 0.2-0.35, respectively. The other parameters for the GA were:

*Population size = 20-40; Number of generations = 200; Crossover rate: 0.8; Elite count: 3-6; Mutation rate: 0.05.*

The 5-fold cross validation techniques were used to estimate the accuracy of each classifier. The grid search algorithm was applied for the stack-vector approach to search for the best parameters (C, γ) for the SVM classifiers. The GA was implemented using the Global Optimization toolbox in Matlab 7.11.1. For the implementation of the SVM classifiers the well-known LIBSVM 3.1 toolkit with Matlab interface was employed [32].

Five land cover classes were identified for classification. These classes were: Crop (CR), Permanent Pastures (PA), Dense Forest (DF, Sparse Forest (SF) and Residential Area (RE). The spectral properties and backscatter signatures of five land cover features in the Landsat 5 TM multi-spectral and multi-date PALSAR images are shown in Figure 5.

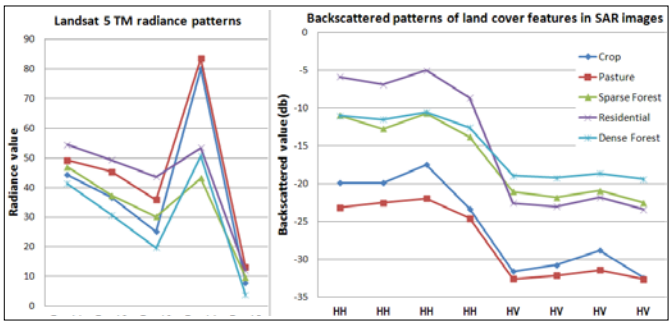


Figure 5 Land cover feature characteristics in Landsat 5 TM (left) and multi-date PALSAR (right) images

The land cover data used for training and validation were derived from visual interpretation with the help of aerial photography, Google Earth images and ground truth data collected from two field surveys conducted on 2 September 2010 and 23 August 2010. The training and test datasets were selected randomly and independently using the Region of Interest (ROI) tool of the ENVI 4.6 software. The sizes of training and testing datasets are given in the below table.

TABLE II CONTENTS OF TRAINING AND TESTING DATASETS

| **Classes** | **Training data** (number of pixels) | **Testing data** (number of pixels) |
|---|---|---|
| Crop | 954 | 1076 |
| Pasture | 956 | 1107 |
| Spare Forest | 961 | 1080 |
| Residential | 313 | 198 |
| Dense Forest | 632 | 569 |

## VI. RESULTS AND DISCUSSIONS

Overall classification accuracies of datasets (consisted of less than 12 features) using the non-feature-selection (stack-vector approach) and SVM classifiers are summarised in Table 3. Results of classification using the proposed feature-selection technique and the non-feature-selection approach for larger datasets are given in Table 5.

The classification results demonstrate the complimentary characteristics and efficiencies from the integration of optical and SAR images. All combined datasets generated from both kinds of data, no matter whether the stack-vector or feature-selection approach was used, produced significant improvements in classification accuracy compared to the original single datasets. In the case of the stack-vector approach, although the Landsat 5 TM gave a high accuracy of 85.21%, it produced extensive confusion between residential and vegetation classes (commission errors were 56.39%). The combination of the original Landsat 5 TM and PALSAR images gave remarkable increases of accuracy, in which the combined use of Landsat TM and four-date PALSAR HH images resulted in classification accuracy of 91.46%, with improvements of 6.25% and 25.60% for Landsat 5 TM and PALSAR HH data respectively, while the commission errors for the residential class were 29.29%. The integration of Landsat 5 TM image with four-date PALSAR dual HH/HV images gave an overall accuracy of 88.93% with an increase of 3.72% and 16.18% over Landsat 5 TM and PALSAR dual HH/HV data, respectively, while the commission errors of the residential class were 39.94%. However, the improvements are more obvious when using the feature selection GA techniques. This integration resulted in an overall accuracy of 92.26% with an increase of 7.05% and 19.51% compared to the single-type datasets, while the residential commission errors were reduced to 19.83%.

The complimentary properties of like- (PALSAR HH) and cross- (PALSAR HV) polarisation images were also clearly highlighted. As can be seen in Table 4, except for the crop where the classification accuracy of PALSAR HH and HV images was rather similar, the like-polarisation gave better accuracy for the residential class (which is dependent on surface scattering), while the cross-polarisation, due to its sensitivity to the volume scattering, provided higher accuracy on vegetation classes: Pasture, Sparse Forest and Dense Forest. Utilisation of both SAR like- and cross-polarised data resulted in noticeable improvements in overall classification accuracy, particularly for the Residential, Sparse Forest and Dense Forest features.

TABLE III CLASSIFICATION ACCURACY OF DIFFERENT DATASETS USING THE SVM CLASSIFIER AND STACK-VECTOR METHODS

| **Datasets** | **Overall Accuracy (%)** |
|---|---|
| Four-date HH images | 65.86 |
| Four-date HH images + $\mathbf{HH}_{AVE}$+$\mathbf{HH}_{CH}$ | 67.92 |
| Four-date HV images | 66.58 |
| Four-date HV images + $\mathbf{HV}_{AVE}$+$\mathbf{HV}_{CH}$ | 68.44 |
| Four-date HH/HV images | 72.75 |
| Four-date HH/HV images + $\mathbf{HH/HV}_{AVE}$ + $\mathbf{HH/HV}_{CH}$ | 73.18 |
| Four-date HH/HV images +$\mathbf{SAR}_{ind}$ | 72.73 |
| Four-date HH images | 65.86 |
| Four-date HH/HV images + $\mathbf{HH/HV}_{AVE}$+$\mathbf{HH/HV}_{CH}$ +$\mathbf{SAR}_{ind}$ | 73.20 |
| L5 TM | 85.21 |

(note: L5 TM = Landsat 5 TM; Four-date HH image = Four-date PALSAR HH polarised image; Four-date HV image = Four-date PALSAR HV polarised image; Four-date HH/HV image = Four-date PALSAR dual HH/HV polarised image)

TABLE IV PRODUCER AND USER ACCURACY (%) OF FOUR-DATE PALSAR HH, HV AND DUAL-POLARISED HH/HV IMAGES

| **Land Cover Classes** | **Four-Date HH** | | **Four-Date HV** | | **Four-Date HH/HV** | |
|---|---|---|---|---|---|---|
| | ***Producer*** | ***User*** | ***Producer*** | ***User*** | ***Producer*** | ***User*** |
| Crop (CR) | 80.11 | 65.80 | 78.07 | 69.19 | 77.79 | 69.35 |
| Permanent Pasture (PA) | 62.87 | 76.48 | 66.03 | 78.69 | 66.58 | 75.51 |
| Sparse Forest (SF) | 59.63 | 70.46 | 70.74 | 67.02 | 78.24 | 77.17 |
| Residential (RE) | 68.18 | 75.84 | 16.16 | 12.90 | 77.78 | 76.24 |
| Dense Forest (DF) | 55.71 | 44.15 | 55.54 | 63.33 | 63.09 | 65.27 |

Incorporation of the PALSAR original images with their additional derived data, such as Average and Temporal Backscattered Change ($\mathbf{SAR}_{CH}$), gave a noticeable increase in classification accuracy. The improvements were 2.06% and 1.86% for the case of four-date PALSAR HH and HV polarised images, respectively. However, in the case of PALSAR dual HH/HV polarised images, there was only a slight increase in accuracy of 0.43%. On the other hand, the SAR index images do not give any improvement compared to the classification of the original PALSAR dual HH/HV images.

Integration of optical and multi-date SAR data with their textural information gave a noticeable increase in the

classification accuracy. As for the non-feature selection approach, the combination of four-date HH images with their textures gave very slight increases in accuracy of 0.81%, 0.92% and 1.34% using the first-order, GLCM and both groups of texture measures, respectively. On the other hand, the increase of 4.41%, 0.79%, and 6.79% in classification accuracy was obtained for the four-date HV images. The improvements for the four-date PALSAR HH/HV images were 4.42%, 1.86% and 6.05%. The application of textural information was also effective for optical images. The integration of Landsat 5 TM images with their textures resulted in increases of 2.48%, 4.37% and 2.95% using the first-order, GLCM and both textural groups respectively.

The efficiency of using textural information was even more impressive when the SVM-GA models were exploited. For example, the improvements for the four-date PALSAR dual HH/HV polarisation increased 6.98%, 2.95% and 6.93%. The combination of Landsat 5 TM images with their corresponding texture measures also gave more significant improvements of 5.53%, 6.00% and 6.23%. In particular, the integration of four-date PALSAR HV polarised images with both types of textures provided an increase of overall classification accuracy of up to 12.95%.

It is worth mentioning that although the four-date PALSAR dual HH/HV images resulted in much lower classification accuracy, 72.75% compared to 85.21% for the Landsat 5 TM image, its combination with textural data and additional images using the SVM-GA approach gave an overall classification accuracy of 81.19%, which is closer to the accuracy of optical images. The comparison of classification improvement by incorporation of textural information using stack-vector and feature-selection GA methods is shown in Figure 6.

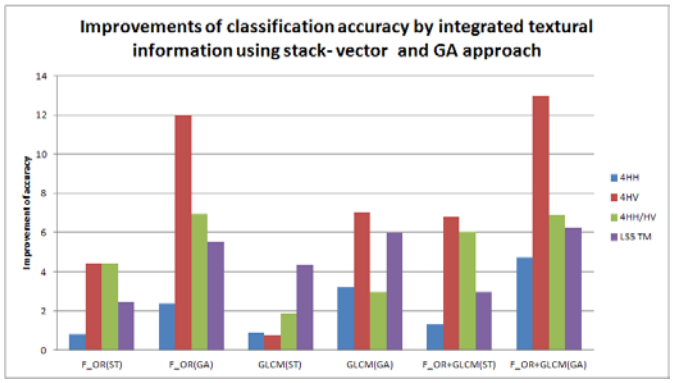


Figure 6 Improvement of accuracy by incorporating textural information with original datasets using stack-vector and feature-selection with GA techniques

(F_OR(ST), GLCM(ST), F_OR+GLCM(ST) and F_OR(GA), GLCM(GA), F_OR+GLCM(GA) represent integration of textures using stack-vector and feature-selection GA methods, respectively)

The SVM-GA approach outperforms the commonly used non-feature-selection approach (Table 5 and Figure 8). In all cases the feature-selection approach gave better results than the non-feature-selection approach. The increase of overall classification accuracy compared to the non-feature-selection methods ranged from 0.87% (four-date PALSAR HH/HV images + its first- and second- order textures) to 7.57% (four-date PALSAR HV images + its first- order textures). It is also important to emphasise that the GA techniques used much less input features than the stack-vector method. A comparison in the number of data input features used for classifications in two approaches is shown in Figure 9.

While in many cases the Hughes phenomenon appeared in the stack-vector method, where the classification accuracy decreased with an increase in the number of input measurements, it is not the case for the SVM-GA method. For example, the integration of the Landsat 5 TM image with PALSAR HH images gave an overall accuracy of 91.46% while the integration of the Landsat 5 TM image with PALSAR dual HH/HV images provided only 88.93% accuracy, which represents a decrease of 2.53%. However, using the SVM-GA model the same integration gave an overall accuracy of 92.28% with only 10 out of a total 13 features used (Figure 7). Similarly, the combination of Landsat 5 TM image with its GLCM textures gave an accuracy of 89.58% but when the Landsat 5 TM image was combined with both first-order and GLCM texture measures the accuracy dropped to 88.16%. Nevertheless, when applying GA techniques the classification accuracy still increased slightly up to 91.44%.

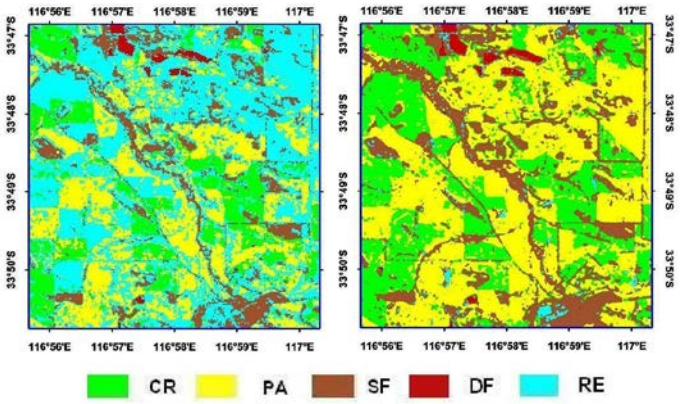


Figure 7 Classification of the Landsat 5 TM image (left) and the integration of Landsat 5 TM & four-date PALSAR HH/HV images (right) using the GA technique

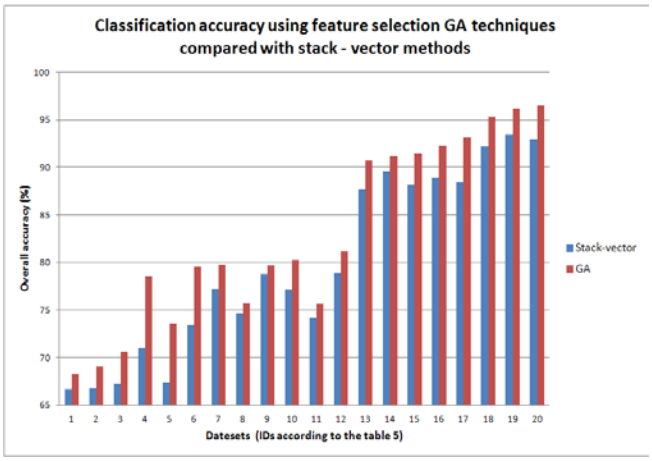


Figure 8 Classification accuracy using feature selection (GA) with non-feature selection (stack-vector) approach

TABLE V COMPARISON OF LAND COVER CLASSIFICATION PERFORMANCE BETWEEN FEATURE-SELECTION AND NON-FEATURE-SELECTION APPROACH; NF=NUMBER OF SELECTED FEATURES

| ID | DATASETS | SVM OVERALL ACCURACY (%) | | | |
|---|---|---|---|---|---|
| | | Stack Vector | | GA | |
| | | *NF* | *Accuracy* | *NF* | *Accuracy* |
| 1 | Four-date HH images + First-order textures | 28 | 66.67 | 8 | 68.24 |
| 2 | Four-date HH images +GLCM textures | 36 | 66.77 | 19 | 69.08 |
| 3 | Four-date HH +First – order & GLCM textures | 60 | 67.20 | 21 | 70.59 |
| 4 | Four-date HV + First-order textures | 28 | 70.99 | 8 | 78.56 |
| 5 | Four-date HV +GLCM textures | 36 | 67.37 | 16 | 73.60 |
| 6 | Four-date HV +First - order & GLCM textures | 60 | 73.37 | 25 | 79.53 |
| 7 | Four-date HH/HV images + First- order textures | 56 | 77.17 | 23 | 79.73 |
| 8 | Four-date HH/HV images +GLCM textures | 72 | 74.61 | 34 | 75.70 |
| 9 | Four-date HH/HV +First -order & GLCM textures | 120 | 78.81 | 51 | 79.68 |
| 10 | Four-date HH/HV images + **HH/HV**$_{AVE}$ + +**HH/HV**$_{CH}$ + **SAR**$_{ind}$ + First-order textures | 64 | 77.15 | 24 | 80.23 |
| 11 | Four-date HH/HV images + **HH/HV**$_{AVE}$ + +**HH/HV**$_{CH}$ + **SAR**$_{ind}$ + GLCM textures | 80 | 74.17 | 36 | 75.65 |
| 12 | Four-date HH/HV images + **HH/HV**$_{AVE}$ + **HH/HV**$_{CH}$ + **SAR**$_{ind}$ + First-order & GLCM textures | 128 | 78.86 | 49 | 81.19 |
| 13 | L5 TM + First-order textures | 29 | 87.69 | 9 | 90.74 |
| 14 | L5 TM + GLCM textures | 37 | 89.58 | 13 | 91.21 |
| 15 | L5 TM + First-order & GLCM textures | 61 | 88.16 | 20 | 91.44 |
| 16 | L5 TM + Four-date HH/HV | 13 | 88.93 | 10 | 92.28 |
| 17 | L5 TM + Four-date HH/HV + **HH/HV**$_{AVE}$ + +**HH/HV**$_{CH}$ + **SAR**$_{ind}$ | 21 | 88.39 | 8 | 93.10 |
| 18 | L5 TM + All of LS5_textures + Four-date HH/HV + **HH/HV**$_{AVE}$+**HH/HV**$_{CH}$ + **SAR**$_{ind}$ | 77 | 92.21 | 32 | 95.24 |
| 19 | L5 TM + All of L5 TM textures + Four-date HH/HV + All of SAR textures | 181 | 93.47 | 81 | 96.20 |
| 20 | L5 TM + All L5 TM textures + Four-date HH/HV + **HH/HV**$_{AVE}$ + **HH/HV**$_{CH}$ + **SAR**$_{ind}$ + All of SAR textures | 189 | 92.98 | 81 | 96.47 |

The highest classification accuracy achieved with the non-feature-selection methods was 93.47% with 181 data input features, while the GA achieved the best accuracy of 96.47% with only 81 selected features.

The use of our proposed fitness function with an additional parameter of average correlation between selected features has improved the overall classification accuracy for 18 out of 20 combined datasets compared to the use of the commonly designed fitness function. There are only two cases in which the classification accuracy decreased very slightly (about 0.1%) while the proposed fitness function was applied. Impacts of the proposed fitness function compared to the commonly used fitness function were illustrated in Figure 10.

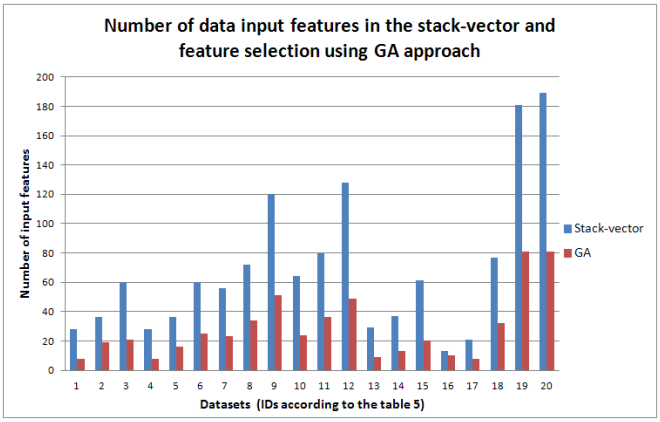


Figure 9 Number of input features in non-feature-selection and feature-selection GA methods

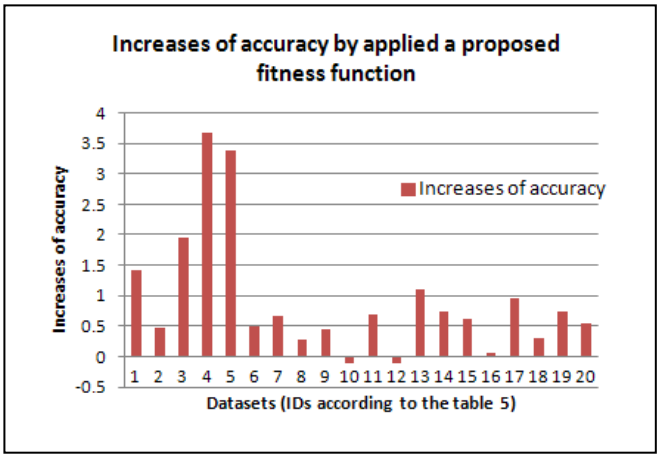


Figure 10 Impacts of the proposed fitness function on overall classification accuracy compared to the commonly used fitness function

## VII. CONCLUSION

A feature selection process based on the Genetic Algorithm and a Support Vector Machine classifier has been proposed and compared with non-feature-selection techniques for classification of multi-source remotely sensed data, including optical, multi-date SAR and textural information. Results of classification of different combined datasets (more than 30 for non-feature-selection and 20 for the feature-selection approach) revealed advantages of multi-source remotely sensed data and SVM-based algorithms for land cover classification. The combination of optical and SAR data often gave higher classification accuracy than any single-type datasets. Incorporation of textural information with either optical or SAR data also resulted in an improvement of accuracy. Feature selection using the SVM-GA approach clearly outperformed the classical stack-vector method. The SVM-GA approach always resulted in better classification accuracy with more significant improvement, and used less data input features compared to the non-feature-selection approach. The highest overall classification accuracy of 96.47% was obtained using the SVM-GA

method for classifying the combined dataset of original Landsat 5 TM, four-date PALSAR dual HH/HV polarised images, all of their textures and additionally derived images with 81 out of 189 data input features selected. Moreover, results of classifications also indicated that the proposed fitness function in this study is more reliable than the commonly used version.

## ACKNOWLEDGMENT

ALOS PALSAR Level 1.1 products were processed by ERSDAC, Japan. Copyright of raw data belongs to METI and JAXA. Landsat 5 TM Level 1T products were processed and provided by the USGS Earth Resources Observation and Science Center.

## REFERENCES

[1] G. Chust, D. Ducrot, and J. Pretus, "Land cover discrimination potential of radar multitemporal series and optical multispectral images in a Mediterranean cultural landscape", International Journal of Remote Sensing, vol. 25(17), 3513–3528 (2004).

[2] S. Erasmi, and A. Twele, "Regional land over mapping in the humid tropics using combined optical and SAR satellite data–a case study from Central Sulawesi, Indonesia", International Journal of Remote Sensing, vol. 30(10), 2465-2478 (2009).

[3] S. H. Kim, and K. S. Lee, "Integration of Landsat ETM+ and Radarsat SAR for land cover classification", Trans Tech Publications, Switzerland, 2005, doi:10.4028/www.scientific.net/KEM.277-279.838.

[4] H. Huang, J. Legarsky, and M. Othman, "Land-cover classification using Radarsat and Landsat imagery for St. Louis, Missouri", Photogrammetric Engineering & Remote Sensing, 73(1), 37–43 (2007).

[5] A. Sheoran, B. Haack, and S. Sawaya, "Land cover/use classification using optical and quad polarization radar imagery", ASPRS 2009 Annual Conference, Baltimore, Maryland, USA, 9-13 March, 2009.

[6] H. T. Chu, and L. Ge, "Synergistic use of multi-temporal ALOS/PALSAR and SPOT multi-spectral satellite imagery for land cover mapping in the Ho Chi Minh city area, Vietnam", Proceeding of IEEE International Geoscience & Remote Sensing Symposyum. (IGARSS), Honolulu, Hawaii, USA, 25-30 July, 2010, pp.1465 – 1468.

[7] J. Soria-Ruiz, Y. Fernadez-Ordonez, and I. H. Woodhouse, "Land-cover classification using radar and optical images: a case study in Central Mexico", International Journal of Remote Sensing, vol. 31(12), 3291–3305 (2010).

[8] B. Haack, and M. Bechdol, "Integrating multisensor data and RADAR texture measures for land cover mapping", Computers & Geosciences, vol. 26, pp.411-421 (2000).

[9] S. Berberoglu, P. J. Curran, C. D. Lloyd, and P. M. Atkinson, "Texture classification of Mediterranean land cover", International Journal of Applied Earth Observation & Geoinformation, vol. 9(3), pp. 322-334 (2007).

[10] D. Lu, M. Batistella, and E. Moran, "Land-cover classification in the Brazilian Amazon with the integration of Landsat ETM+ and Radarsat data", International Journal of Remote Sensing, vol. 28(24), 5447—5459 (2007).

[11] A. N. Tassetti, E.N. Malinverni, and M. Hahn, "Texture analysis to improve supervised classification in IKONOS imagery", ISPRS TC VII Symposium – 100 Years ISPRS, Vienna, Austria, 5-7 July, 2010, IAPRS, Vol. XXXVIII, Part 7A, pp. 245-250.

[12] C. A. Coburn, and A. C. Roberts, "A multiscale texture analysis procedure for improved forest stand classification", International Journal of Remote Sensing, vol. 25(20), 4287–4308 (2004).

[13] Y. Zhang, L. Chen, and B. Yu, "Multi-scale texture analysis for urban land use/cover classification using high spatial resolution satellite data", Geoinformatics 2007: Remotely Sensed Data and Information, Proc. of SPIE Vol. 6752, 67523G, doi:10.1117/12.761232.

[14] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers". IEEE Trans. Inf. Theory, vol. 14(1), pp. 55–63 (1968).

[15] D. Lu, and Q. Weng, "A survey of image lassification methods and techniques for improving classification performance", International Journal of Remote Sensing, vol. 28(5), pp. 823-870 (2007).

[16] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification", International Journal of Remote Sensing, vol. 23(4), 725–749 (2002).

[17] B. Waske, and M. Braun, "Classifier ensembles for land cover mapping with multi-spectral SAR imagery", International Journal of Photogrammetry and Remote Sensing vol. 5, pp. 450-457 (2009).

[18] B. Waske, and J. A. Benediktsson, "Fusion of Support Vector Machines for classification of multisensor data", IEEE Transactions on Geosciences & Remote Sensing, vol. 45(12), pp. 3858–3866 (2007).

[19] H. Z. M. Shafri, and F. S. H. Ramle, "A comparison of Support Vector Machine and Decision Tree Classifications Using Satellite Data of Langkawi Island", Information Technology Journal, vol. 8 (1): pp. 64-70 (2009).

[20] M. Pal, and P. M. Mather, "Support Vector Machines for classification in remote sensing", International Journal of Remote Sensing, vol. 26, pp. 1007–1011 (2005).

[21] T. Kavzoglu, and I. Colkesen, "A kernel functions analysis for support vector machines for land cover classification", International Journal of Applied Earth Observation & Geoinformation, vol. 11(5), pp. 352–359 (2009).

[22] C. W. Hsu, C. C. Chang, and C. J. Lin, "A Practical Guide to Support Vector Classification", Department of Computer Science, National Taiwan University (2009). Available from http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

[23] C. L. Huang, and C. J. Wang, "A GA-based feature selection and parameters optimization for support vector machines", Expert Systems with Applications, vol. 31, pp. 231-240 (2006).

[24] L. Zhuo, J. Zheng, F. Wang, X. Li, B. Ai, and J. Qian, "A Genetic Algorithm based wrapper feature selection method for classification of hyperspectral images using support vector machine", The International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences. Vol. XXXVII. Part B7. Beijing, pp. 397-402.

[25] M. Zhou, J. Shu, and Z. Chen, "Classification of hyperspetral remote sensing image based on Genetic Algorithm and SVM", *Remote Sensing & Modeling of Ecosystems for Sustainability VII*, Proc. of SPIE, vol. 7809, 78090A (2010).

[26] R. Osuna, Pattern Analysis, Sequential Feature Selection, Lecture note. Texas A&M University, 2005.

[27] M. Pal, "Support vector machine-based feature selection for land cover classification: a case study with DAIS

hyperspectral data", International Journal of Remote Sensing, vol. 27(14), pp. 2877- 2894 (2006).

[28] S. Osowski, R. Siroic, T. Markiewicz, and K. Siwek, "Application of Support Vector Machine and Genetic Algorithm for improved blood cell recognition", IEEE Transactions on Instrumentation and Measurement, 58(7), 2159–2168 (2009).

[29] V. N. Vapnik, The Nature of Statistical Learning Theory, New York : Springer, 2000.

[30] A. Lopes, R. Touzi, and E. Nezry, "Adaptive Speckle Filters and Scene Heterogeneity", IEEE Transactions on Geoscience & Remote Sensing, vol. 28(6), pp. 992-1000 (1990).

[31] M. Hall-Beyer, GLCM Texture Tutorial; Version: 2.10, University of Calgary, 2007. Available from: http://www.fp.ucalgary.ca/mhallbey/tutorial.htm.

[32] C. H. Chang, and C. J. Lin, "LIBSVM: a library for support vector machines", ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27 (2011).

[33] J. R. Jensen, Introduction to Digital Image Processing: A Remote Sensing Perspective, 4th edition, Prentice Hall, 2004.